

Exhibit A

UNITED STATES DISTRICT COURT
EASTERN DISTRICT OF VIRGINIA
ALEXANDRIA DIVISION

UNITED STATES, *et al.*,

Plaintiffs,

vs.

GOOGLE LLC,

Defendant.

No: 1:23-cv-00108-LMB-JFA

**DECLARATION OF BRADEN SHEFF IN SUPPORT OF
GOOGLE LLC'S MEMORANDUM IN RESPONSE TO PLAINTIFFS'
MOTION FOR INTERIM MODIFICATIONS TO THE DISCOVERY PLAN**

I, Braden Sheff, pursuant to 28 U.S.C. § 1746, hereby declare as follows:

1. I am a Manager, Discovery Project Management, at Defendant Google LLC ("Google"). I have held this title since 2020 and have been employed by Google since 2014. I submit this declaration in support of Google's Memorandum in Response to Plaintiffs' Motion For Interim Modifications To The Discovery Plan Due To Google's Non-Compliance With Discovery Deadlines in *United States v. Google LLC*, No. 1:23-cv-00108-LMB-JFA (E.D. Va.) (the "Motion").

2. The contents of this declaration are true and correct to the best of my knowledge, information, and belief, and are based on my personal knowledge of Google's discovery processes, and a reasonable inquiry into the issues arising in this case, informed by my conversations with colleagues at Google. I am competent to testify to the matters set forth herein.

3. In response to a regulatory investigation or other anticipated litigation, it is Google's usual practice to collect the documents of employees with relevant knowledge of the investigation or litigation ("custodians") and house them in a database called an ingestion site. For a typical litigation, Google's collections require only one ingestion site.

4. The database structure in this matter is atypical for two reasons. First, this case follows on multiple investigations, including the Texas Attorney General's investigation of Google's ad tech business (announced in September 2019) and the DOJ's omnibus investigation of Google (which began in 2019 and ultimately led to multiple litigations) (together, the "Investigations"). Second, the volume of information collected in this matter was exceptionally high. Six ingestion sites were used to collect and house these documents, which amounted to approximately 80 terabytes of data and spanned years and hundreds of custodians. One of these ingestion sites related to the Texas Attorney General's investigation ("Texas AG Investigation Ingestion Site" and, together with the DOJ Ingestion Sites, the "Other Ingestion Sites"). The other five ingestion sites ("DOJ Ingestion Sites") flowed from the DOJ's investigation.

5. When search terms are used, Google's typical practice is to export documents that hit on search terms from an ingestion site to a third-party ediscovery vendor so the documents can be processed and prepared for review.

6. Consistent with that practice, in this matter, Google applied DOJ's investigation-phase search terms to documents collected in the DOJ Ingestion Sites, exported those documents for review, and ultimately produced nearly 3 million documents to the Department of Justice before this lawsuit was filed (including certain documents originally produced to the Texas Attorney General from the Texas AG Investigation Ingestion Site).

7. When discovery opened in the related MDL litigation in the Southern District of New York and this lawsuit was filed, Google agreed to refresh its investigative production to the Department of Justice. To do so, Google identified a set of 119 ad tech custodians and collected their documents through March 27, 2023.¹ Due to the size of this collection and limitations of the existing ingestion sites, Google stored this refreshed collection in a new ingestion site (the “Ad Tech Litigation Ingestion Site”), which holds over 54 million documents, comprising over 21 terabytes of data.

8. MDL Plaintiffs served 301 Requests for Production and DOJ served 76. During the spring and summer of 2023, Google negotiated search terms and custodians with Plaintiffs in this litigation and MDL Plaintiffs. During these negotiations, Google agreed to collect and produce documents from additional custodians, bringing the total number of custodians from 119 up to 172 (the “Ad Tech Custodians”). As custodians were added, their documents were collected for the decade-long time period negotiated with Plaintiffs in this case (“Agreed Time Period”)² and stored in the Ad Tech Litigation Ingestion Site. Certain additional non-custodial shared drives to which Ad Tech Custodians had access were collected into a second litigation ingestion site (collectively with the Ad Tech Litigation Ingestion Site, the “Litigation Ingestion Sites”).

¹ For most custodians this “refresh” collection covered the several years between the last date on which their documents had been collected for the Investigations and March 27, 2023. However, for some custodians the refresh collections also covered time periods before October 20, 2018 because the DOJ had not requested documents for the January 1, 2013 through October 20, 2018 time period for those custodians during its investigation. If a custodian’s documents for a particular date range had previously been collected during the Investigations, Google generally did not recollect those documents for this “refresh” collection.

² For most Ad Tech Custodians, the Agreed Time Period is January 1, 2013 through March 27, 2023. However, for 17 custodians, the Agreed Time Period begins on January 1, 2009 (and, for one of these 17, ends on June 30, 2014), and for 20 other custodians, the Agreed Time Period is October 1, 2020 through March 27, 2023.

9. By June 2023, and at the insistence of Plaintiffs in this litigation and the related MDL litigation, Google agreed to run over 200 search strings (the “Litigation Search Terms”) across the documents collected from the 172 Ad Tech Custodians.

10. The Litigation Search Terms were inadvertently run only on the Litigation Ingestion Sites, and not on the DOJ Ingestion Sites or the Texas AG Ingestion Site. As a result, fewer hit counts were returned, suggesting these terms provided a voluminous but reasonable population to begin processing for review, and only documents from the Litigation Ingestion Sites were exported to Google’s ediscovery vendor for processing for review.

11. This was the result of inadvertent human error. The team member at Google who ran the Litigation Search Terms across the files in the Litigation Ingestion Sites had not worked on the case during the pre-suit investigation phase and did not know of the existence of the many overlapping collections housed in the DOJ Ingestion Sites and the Texas AG Investigation Ingestion Site. As a result, the team member mistakenly believed that the Litigation Ingestion Sites were the only sites in which to run the Litigation Search Terms. Due to the volume of the documents captured by the search terms (5.87 million after deduplication), no anomaly was evident. Until this issue came to light, others on the team did not realize that knowledge of the overlapping ingestion sites was not held by each team member.

12. Google became aware of a potential issue with the document productions in this case in mid-August 2023. On or about August 15, the number of search term hit counts for a single Google custodian noticed for deposition appeared low in contrast to the overall population of produced documents. Google investigated these document counts and determined on August 16 that the bulk of that custodian’s documents were housed in at least one Other Ingestion Site over which the Litigation Search Terms had not been run.

13. Google immediately began investigating whether other custodians beyond that single deponent could be impacted.

14. Scoping this issue has been a time intensive process, primarily because of the machine time required for Google to export search term hits from ingestion sites to its external ediscovery vendor, and then for the vendor to process and deduplicate the documents against documents already included in the review platform. Given the length of the Agreed Time Period, and the multiple collections arising from separate investigations and litigations across 172 custodians, there is significant overlap between the date ranges of documents stored in the Litigation Ingestion Sites and the Other Ingestion Sites. Raw numbers of documents exported to the review platform therefore do not meaningfully indicate the number of documents requiring review. Deduplication is a critical step in assessing scope, and it cannot be completed until all of the files are exported, processed, and uploaded to the review platform.

15. Google immediately prioritized exporting the files of noticed deponents from the DOJ Ingestion Sites to the review platform. Between August 17 and August 19, Google ran all Litigation Search Terms over the deponents' documents in the DOJ Ingestion Sites for the Agreed Time Period and began exporting these documents for review.³

16. Google continued investigating the scope of the issue over the weekend of August 19 and 20 and throughout the week of August 21, working daily with its ediscovery vendor to ascertain how many unique documents of deponents were inadvertently not exported.

17. On August 28, when it had become clear that Google would not have a complete sense of the scope of the issue in the immediate future, Google notified Plaintiffs of a document production issue that could impact document productions for deponents.

³ These searches did not cover the Texas AG Investigation Ingestion Site described in Paragraph 21.

18. As of August 30, Google was still working to substantially complete production of custodial documents from all upcoming deponents in advance of the scheduled depositions.

19. However, the scope of the issue continued to grow as Google investigated the impact of the issue on other custodians who were not scheduled to be deposed. At the same time, Google did not know and could not reliably estimate the number of documents that would need to be reviewed because of the significant time to export, process, and deduplicate documents. Despite this uncertainty, on the morning of September 1, Google informed Plaintiffs it would likely have a substantial number of additional documents to produce and that the issue could impact the fact discovery deadline. Google informed the Court of this on the same day.

20. Google continued to export documents to its outside ediscovery vendor for processing and deduplication throughout the week of September 4. Google provided regular updates to Plaintiffs during that week regarding scope and timing. Google also retained an additional external ediscovery vendor in order to review documents as quickly and efficiently as possible. Google has produced approximately 52,000 documents to Plaintiffs after reviewing documents exported from the DOJ Ingestion Sites.

21. Late on the evening of September 8, as part of its ongoing scoping, Google discovered that another ingestion site (the Texas AG Investigation Ingestion Site) contained documents from 24 custodians for some date ranges within the Agreed Time Period that were not collected in either the Litigation Ingestion Sites or the DOJ Ingestion Sites, and updated Plaintiffs the following business day. Until September 8, Google believed that all of the Ad Tech Custodians' documents for the Agreed Time Period were housed in the Litigation Ingestion Sites and DOJ Ingestion sites.

22. Google has cross-checked the date ranges of each of the collections for each of the 172 custodians against the Agreed Time Period to confirm that, for each of the custodians, the Litigation Search Terms have been run across documents for the entire Agreed Time Period (as far back as 2009 and extending through March of this year) and those documents have been exported, or are in the process of exporting, to the review platform. Since the issue was first discovered, nearly 16 million documents have been, or are in the process of being, exported. These documents need to be deduplicated against documents already in the review platform; responsive, non-privileged documents will be produced.

23. Google has been exporting batches of documents from the DOJ Ingestion Sites to its outside ediscovery vendor for processing, promotion to the review platform and deduplication as quickly as they could be processed and exported since August 17, and it will continue to do so until all Litigation Search Term hits for all custodians over the entire Agreed Time Period have been exported. Since September 8, the same process has been underway for the Texas AG Investigation Ingestion Site. Of these exports, so far nearly 6.3 million documents are duplicates and roughly 3.1 million documents will be added to the review population. Approximately 6.3 million documents remain to be processed and deduplicated.

24. Google's outside ediscovery vendor has been releasing documents from the DOJ Ingestion Sites for review on a rolling basis since August 23. Since August 28, Google has been producing responsive, non-privileged documents from this set on a rolling basis.

25. As explained above, the issue here arose from an inadvertent error when fewer than all ingestion sites were selected to run the Litigation Search Terms in. Based on my review of correspondence, the issue in this case is distinct from a discovery issue that arose in *United States v. Google*, 1:20-cv-03010 (D.D.C.) earlier this year—since remedied—in which Google

notified the Department of Justice that certain files located in Google Drive and shared/team drive folders had not been collected in that case due to errors. Google collected, reviewed, and produced those file types in this case.

I declare under penalty of perjury that the foregoing is true and correct.

Executed on the 13 day of September, 2023, in Austin, Texas.

/s/ Braden Sheff

Braden Sheff